
Classificação

Gabriel Fernandes Silva

Pré-Processamento

1420 IES

1335 ativas

85 extintas

#	Column	Non-Null Count		Dtype
0	Codigo da IES	1420	non-null	int64
1	Fim Lucrativo	1420	non-null	int64
2	Situacao	1420	non-null	int64
3	DIAS	1420	non-null	int64
4	Cursos Sem Ato 5 anos	1420	non-null	float64
5	Ultimo CI	1420	non-null	float64
6	Ultimo IGC	1420	non-null	float64
7	EAD_17	1420	non-null	int64
8	Variacao Matricula 16/17	1420	non-null	float64
9	Matriculas 17	1420	non-null	int64
10	% FIES	1420	non-null	float64
11	CURSOS	1420	non-null	int64
12	Saldo 2017	1420	non-null	float64
13	Variacao do Saldo 16/17	1420	non-null	float64

Modelo 1

KNN

Separação dos Dados

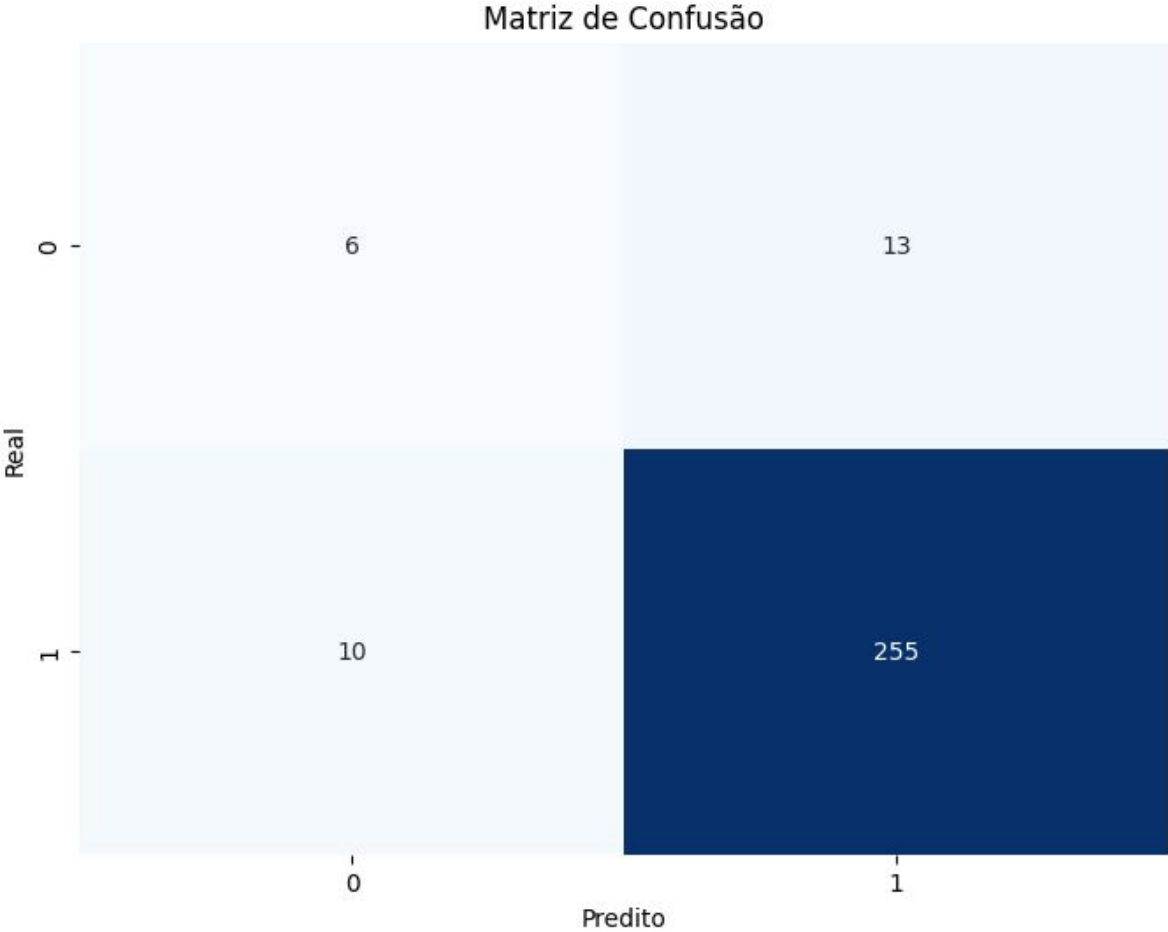
- 80% treino
- 20% teste
- `n_neighbors = 3`

—

Métricas

	P	R	F1	S
Extinta	38%	32%	34%	19
Ativas	95%	96%	96%	265

Acurácia: 92%



Modelo 2

Árvore de Decisão



Separação dos Dados

- 80% treino
- 20% teste
- Sem configuração de hiperparâmetros

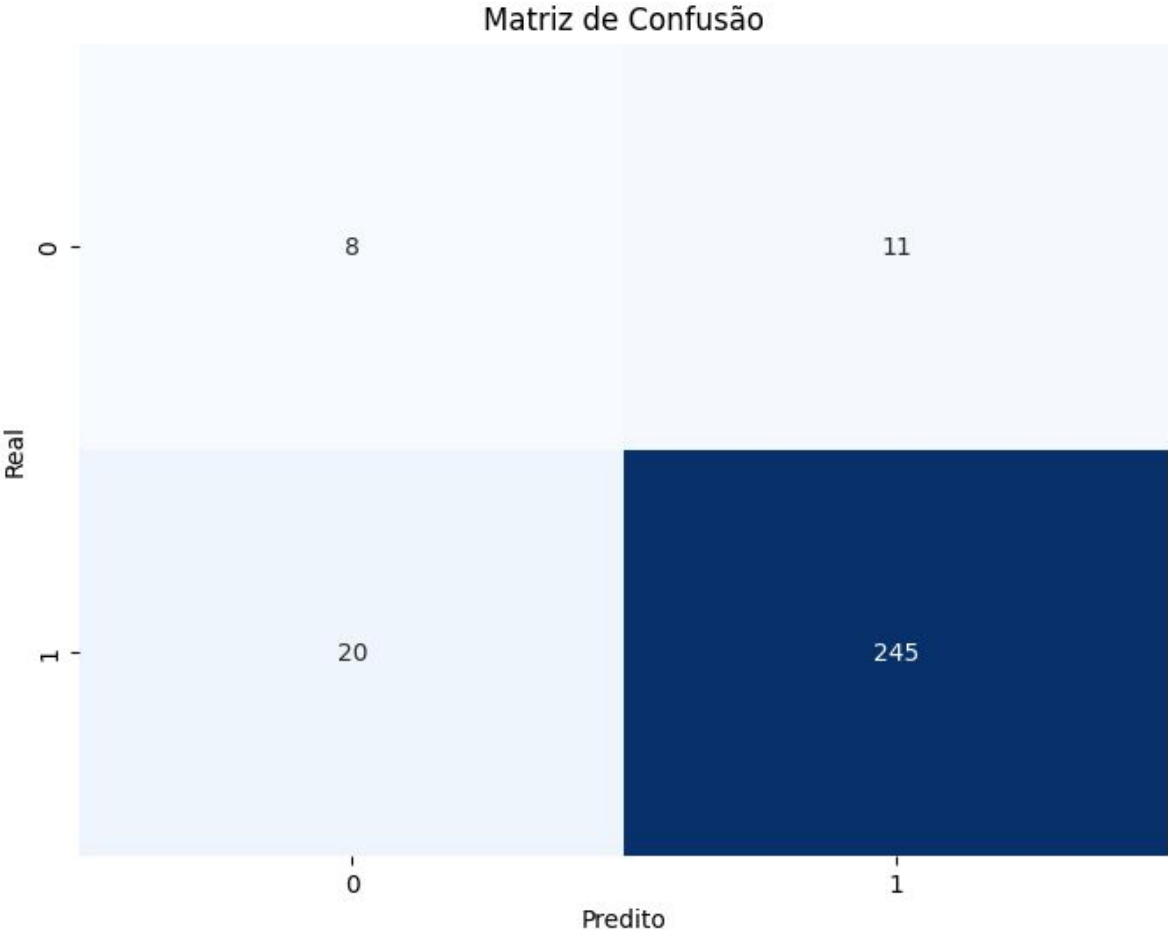


—

Métricas

	P	R	F1	S
Extinta	29%	42%	34%	19
Ativas	96%	92%	94%	265

Acurácia: 89%



Visualização da árvore

Separação dos Dados

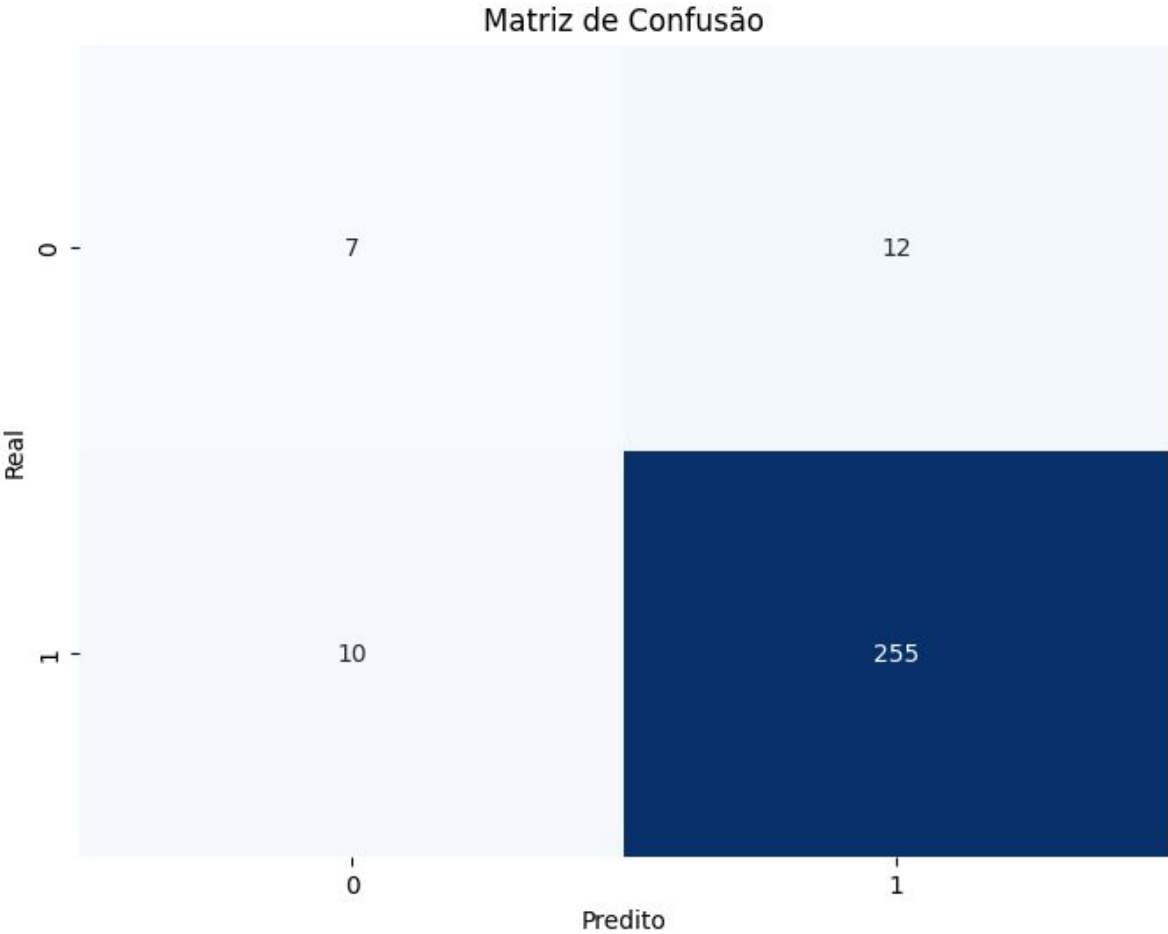
- 80% treino
 - 20% teste
 - `max_depth = 5`
 - `min_samples_split = 5`
 - `min_samples_leaf = 5`
-

—

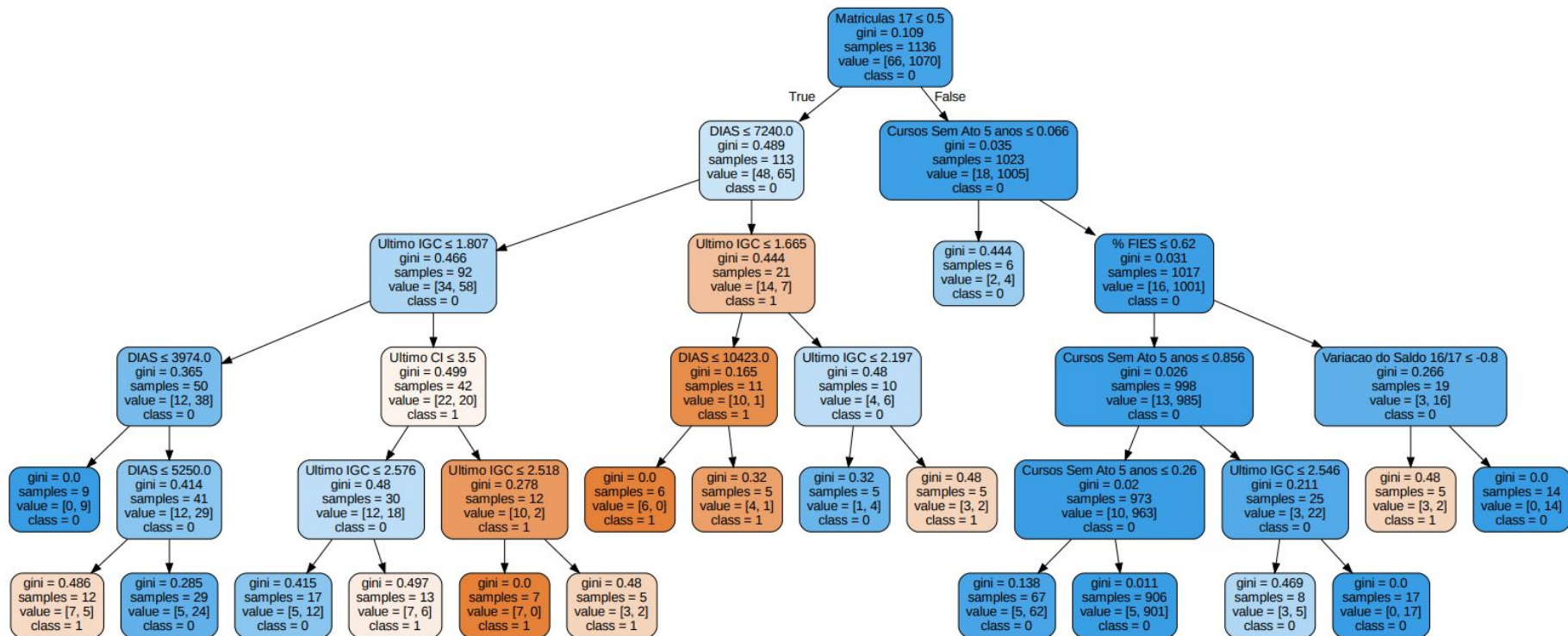
Métricas

	P	R	F1	S
Extinta	41%	37%	39%	19
Ativas	96%	96%	96%	265

Acurácia: 92%



Visualização da árvore



Conclusões

Comparação

KNN

- Apresentou uma sensibilidade mais baixa em relação a instituições extintas

Árvore de Decisão

- O modelo sem ajustes apresentou maior sensibilidade a instituições extintas
-

Melhor Modelo

- Árvore de Decisão se mostrou mais interessante para o problema de identificar instituições extintas

Principais Indicadores:

- Curso Sem Ato 5 anos
 - DIAS
 - Ultimo IGC
-

Obrigado!
